



Time-Scale Block Bootstrap tests for non Gaussian finite variance self-similar processes with stationary increments

Herwig Wendt, Patrice Abry

► To cite this version:

Herwig Wendt, Patrice Abry. Time-Scale Block Bootstrap tests for non Gaussian finite variance self-similar processes with stationary increments. IEEE Statistical Signal Processing Workshop 2007 (SSP'07), 2007, Madison, Wisconsin, United States. ensl-00160708

HAL Id: ensl-00160708

<https://hal-ens-lyon.archives-ouvertes.fr/ensl-00160708>

Submitted on 6 Jul 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TIME-SCALE BLOCK BOOTSTRAP TESTS FOR NON GAUSSIAN FINITE VARIANCE SELF-SIMILAR PROCESSES WITH STATIONARY INCREMENTS

Herwig Wendt, Patrice Abry

CNRS UMR 5672, Physics Dept., Ecole Normale Supérieure de Lyon, France.
herwig.wendt@ens-lyon.fr, patrice.abry@ens-lyon.fr

ABSTRACT

Scaling analysis is nowadays becoming a standard tool in statistical signal processing. It mostly consists of estimating scaling attributes which in turns are involved in standard tasks such as detection, identification or classification. Recently, we proposed that confidence interval or hypothesis test design for scaling analysis could be based on non parametric bootstrap approaches. We showed that such procedures are efficient to decide whether data are better modeled with Gaussian fractional Brownian motion or with multifractal processes. In the present contribution, we investigate the relevance of such bootstrap procedures to discriminate between non Gaussian finite variance self similar processes with stationary increments (such as Rosenblatt process) and multifractal processes. To do so, we introduce a new joint time-scale block based bootstrap scheme and make use of the most recent scaling analysis tools, based on wavelet leaders.

Index Terms— Scaling Analysis, Self similar process, Rosenblatt process, Wavelet Leader, Non Parametric Bootstrap, Confidence Intervals, Hypothesis Tests.

1. INTRODUCTION

Scaling analysis. Scaling analysis is becoming a standard tool in empirical time series analysis. Essentially, it can be summarized as follows. Let $X(t)$ denote the time series to be analyzed and $T_X(a, t)$ some multiresolution quantities, that depend jointly on an analysis scale a and time t . Wavelet coefficients are typical examples of such $T_X(a, t)$. X is said to possess scaling properties when its structure functions,

$$S(a, q) = 1/n_a \sum_{k=1}^{n_a} |T_X(a, t)|^q, \quad (1)$$

behave as power laws with respect to the analysis scale a , for given ranges of analysis scales $a \in [a_m, a_M]$ ($a_M/a_m \gg 1$) and statistical orders q :

$$S(a, q) \sim F_q |a|^{\zeta(q)}. \quad (2)$$

Scaling analysis consists of estimating the scaling exponents, $\zeta(q)$, which, in turns, are involved in standard signal processing tasks such as detection, identification or classification. Scaling exponents can be related to the multifractal spectrum and properties of X . This will not be discussed here and the reader is referred to e.g., [1, 2].

Scaling modeling. A key practical issue consists of deciding which scaling model better fits the data. There exist two major classes of scaling processes: Self similar processes with stationary increments

(H -sssi), whose most popular member is fractional Brownian motion (FBM), and multifractal (MF) processes. In most of the literature related to practical analysis of scaling (or multifractal) properties of empirical data, practitioners are trying to assess whether the data they analyze are better described by one class of processes or by the other. Such a systematic concern corresponds to both theoretical and practical preoccupations:

First, while self similar processes with stationary increments are deeply tied to additive random walks, the most prominent examples of multifractal processes (such as the celebrated Mandelbrot's cascades) are based on recursive multiplicative constructions. Thus, one can naturally imagine that the physical or biological mechanisms underlying the data are likely to be significantly different for data that can be associated to additive or to multiplicative structures.

Second, while the scaling properties of self similar processes are entirely described by a single parameter H , referred to as the self-similarity parameter, those of multifractal processes involve a whole collection of such parameters. Since the goal of practitioners consists of relating these scaling parameters to *physical* parameters controlling the data or the experiment, there is a need to decide on the actual number of independent scaling parameters to be involved.

Third, while practitioners often prefer multifractal models whose increased number of parameters brings extra degrees of freedom and hence better fits of the data, this additional flexibility is achieved at the price of significantly increased practical difficulties. Indeed, while self similar processes correspond to a well-defined fully parametric class of stochastic additive processes mostly depending on a single scaling exponent, multifractal processes fall into a much broader and less well-defined class of multiplicative processes involving a larger number of scaling parameters, for which the derivation of estimation or hypothesis test procedures and the assessment of their statistical performance is much more delicate.

These arguments (addition vs. multiplication, single vs. many parameters, significant extra technical and practical difficulties in estimation and modeling) underline the need for practical procedures helping in deciding whether MF processes are needed to model data, or whether they are satisfactorily well modeled by self-similar ones.

Self similar versus multifractal processes. In the remainder of the text, we discuss only processes with finite variance. For H -sssi processes, the scaling exponents take the specific form $\zeta(q) = qH$. In contrast, for most MF processes of interest, $\zeta(q)$ is a non linear (concave) function of q . We will use this property as a definition for MF processes. Note that such a definition is not strictly correct (cf. [1, 2] for a mathematically correct definition). It is nevertheless sufficient for the context of this contribution and will be seen as a practical or operational definition for MF processes. Therefore, discriminating between these two classes of models mostly amounts to deciding on the linearity of $\zeta(q)$ with respect to q . To this end, the scaling

The authors thank an anonymous reviewer for fruitful comments.

exponents can be fruitfully expanded as $\zeta(q) = \sum_{p \geq 1} c_p q^p / p!$. Indeed, for H -sssi processes, $c_1 = H$ and $c_p \equiv 0, \forall p \geq 2$, while for MF processes $c_2 \neq 0$. This explains why a key issue in scaling analysis consists of accurately estimating the parameters c_2 and c_3 and/or of performing hypothesis tests such as $c_p = 0$ vs. $c_p \neq 0$ (for $p = 2, 3$). Because scaling processes are characterized by intricate properties (long range dependency, non Gaussianity), it is difficult, and so far unachieved, to derive analytically, even asymptotically, the statistical performance of the estimation and test procedures used in scaling analysis. Hence, the lack of confidence intervals for estimates or performance for tests constitutes a major practical limitation in empirical scaling analysis.

Time-scale block Bootstrap. To overcome such difficulties, we studied, in a previous work (cf. [3]), how non parametric bootstrap based approaches can provide the practitioners with both accurate confidence intervals for the estimation of scaling attributes, $\zeta(q)$ and c_p , and meaningful hypothesis tests. To do so, we made use of both wavelet coefficients and wavelet Leaders for scaling analysis [1] and applied (time-block) bootstrap schemes in these wavelet domains. This enabled us to derive efficient procedures for discriminating between (Gaussian) FBM and non Gaussian MF processes [3,4]. In the present contribution, we further investigate the performance of bootstrap based approaches when applied to non Gaussian finite variance self similar processes with stationary increments, and compare them to those obtained from Gaussian FBM. The Rosenblatt process [5] (ROS) will be used as a very example for such processes.

Wavelet coefficients are usually characterized by a joint time and scale dependence structure, which theoretically needs to be taken into account in bootstrap procedures. Our previous scheme consisted of time-block bootstrap for each given scale $a = 2^j$ independently, hence accounting for the time dependence structure of the wavelet coefficients but neglecting the scale one. For Gaussian FBM, it has been shown (see e.g., [6]) that correlation and hence dependence between and amongst scales decreases fast. For non Gaussian H -sssi, while inter-scale correlations decrease as that of Gaussian FBM, inter-scale dependence are likely to remain larger. This may decrease the performance of time-only block bootstrap procedures.

To correct for this weakness, we introduce, in the present contribution, a wavelet domain joint time-scale block bootstrap procedure. We study the performance of such bootstrap schemes and show that they provide practitioners with satisfactory confidence intervals for scaling attributes and powerful hypothesis tests aiming at discriminating between potentially non Gaussian H -sssi and MF processes.

2. FINITE VARIANCE SELF SIMILAR PROCESSES

Self similar processes with stationary increments, and self-similarity parameter H , $0 < H < 1$, referred to as H -sssi processes, satisfy:

$$\{X(t)\}_{t \in \mathbb{R}} \stackrel{fdd}{=} \{a^H X(t/a)\}_{t \in \mathbb{R}}, \forall a > 0, \quad (3)$$

$$\{X(t+a) - X(t)\}_{t \in \mathbb{R}} \stackrel{fdd}{=} \{X(a)\}_{a \in \mathbb{R}}, \quad (4)$$

where $\stackrel{fdd}{=}$ stands for the equality of all finite dimensional distributions [7]. FBM is the only Gaussian H -sssi process [7]. There also exist finite variance non Gaussian H -sssi processes, the most well known being the *Rosenblatt process* (ROS) [5]. The Rosenblatt process, with $H \in (1/2, 1)$, is defined as

$$Z_H(t) = k_H \int_{\mathbb{R}^2} \left\{ \int_0^t (s-u)_+^{\frac{H}{2}-1} (s-v)_+^{\frac{H}{2}-1} ds \right\} dB(u) dB(v), \quad (5)$$

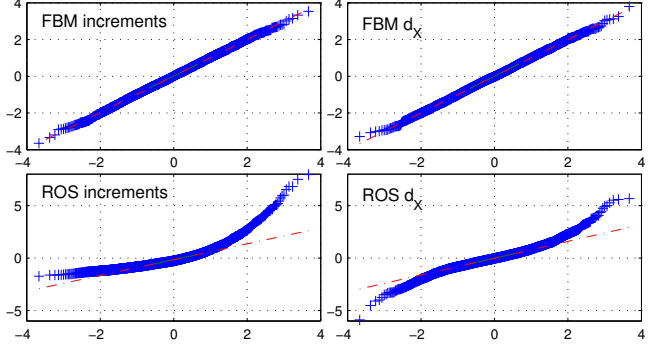


Fig. 1: Quantile-quantile plots of normalized empirical distributions of increments (left) and wavelet coefficients (right) versus standard normal (quantiles given by abscissa) for FBM (top) and Rosenblatt (bottom). Whereas the distributions for FBM are Gaussian, the distributions for ROS are highly non Gaussian and skewed.

where k_H is a normalizing constant (e.g. such that $\mathbb{E}Z_H(1)^2 = 1$), $\int_{\mathbb{R}^2}$ denotes the double Wiener-Itô integral, $x_+ = \max\{x, 0\}$ for $x \in \mathbb{R}$ and $B(t)$, $t \in \mathbb{R}$, is a standard Brownian motion. For further technical details on the definition, properties and numerical simulation procedures, the reader can consult e.g., [8] and references therein. Its two major properties of interest are as follows: ROS has exactly the same covariance function as FBM but more complex higher order dependence. Also, ROS has a non Gaussian highly skewed marginal distribution (as illustrated from numerical simulations in Fig. 1). Such characteristics (complex dependence and non Gaussianity) are likely to significantly impair scaling analysis and the capabilities of bootstrap to provide relevant confidence limits and/or hypothesis tests. This is precisely the question under analysis in the present contribution.

3. SCALING ANALYSIS

Multiresolution quantities. Scaling analysis is nowadays commonly performed using wavelet coefficients as multiresolution quantities. However, it has been shown and explained in [1] that wavelet Leaders can interestingly replace wavelet coefficients, yielding a more relevant, accurate and general multifractal analysis. We will apply bootstrap procedures to both wavelet coefficient and Leaders and compare their performance.

The discrete wavelet transform of X is defined as: $d_X(j, k) = \int_{\mathbb{R}} X(t) 2^{-j} \psi_0(2^{-j}t - k) dt$, where $\psi_0(t)$ is an elementary function, referred to as the *mother-wavelet* and characterized by its *number of vanishing moments* $N_\psi \geq 1: \forall k = 0, 1, \dots, N_\psi - 1, \int_{\mathbb{R}} t^k \psi_0(t) dt \equiv 0$ and $\int_{\mathbb{R}} t^{N_\psi} \psi_0(t) dt \neq 0$.

It is moreover assumed that ψ_0 has a compact time support and that the collection $\{\psi_{j,k}(t) = 2^{-j/2} \psi_0(2^{-j}t - k), j \in \mathbb{Z}, k \in \mathbb{Z}\}$ form an orthonormal basis of $L^2(\mathbb{R})$. Let us define dyadic intervals as $\lambda = \lambda_{j,k} = [k2^j, (k+1)2^j)$, and let 3λ denote the union of the interval λ with its 2 adjacent dyadic intervals: $3\lambda_{j,k} = \lambda_{j,k-1} \cup \lambda_{j,k} \cup \lambda_{j,k+1}$. Then, following [1], wavelet Leaders are defined as: $L_X(j, k) \equiv L_\lambda = \sup_{\lambda' \subset 3\lambda} |d_{X,\lambda'}|$.

Scaling properties. Both for H -sssi and MF processes, it has been

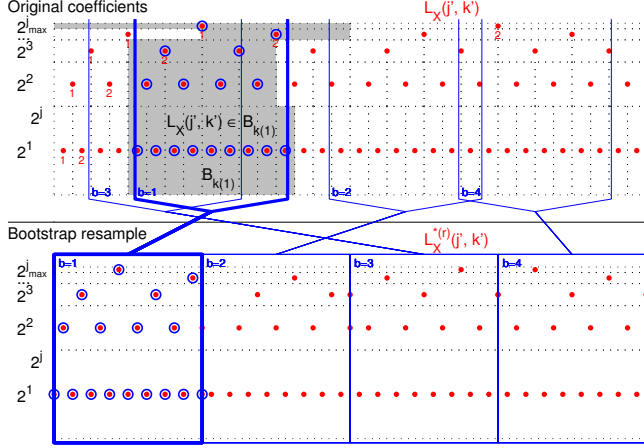


Fig. 2: Illustration of time-scale block bootstrap for $n = 2^6$ and $l = 2^3$. On top, the set of original coefficients (red dots) in the time-scale plane. The coefficients marked by blue circles belong to the time-scale block $B_{k(1)}$. The bootstrap resample $L_X^{*(r)}$ (bottom) consists of the coefficients that lie within the $B = \lceil \frac{n}{2l} \rceil = 4$ time-scale blocks $B_{k(b)}$, $b = 1, \dots, B$, drawn independently and with replacement from all available overlapping circular time-scale blocks.

shown that the scaling properties of X can be expressed as:

$$S(2^j, q) = F_q 2^{j\zeta(q)}, \quad (6)$$

$$C(2^j, p) = c_p^0 + c_p \ln 2^j, \quad (7)$$

where $S(2^j, q)$ and $C(2^j, p)$ denote, respectively, the structure functions as defined in Eq. (1), and the cumulant of order p of the logarithm of the multiresolution quantities (i.e., of $\ln L_X(j, k)$ or $\ln |d_X(j, k)|$).

Estimation. Eqs. (6) and (7) suggest that the estimation of the $\zeta(q)$ and c_p can be performed by means of linear regressions:

$$\hat{\zeta}(q) = \sum_{j=j_1}^{j_2} w_j \log_2 S(2^j, q), \quad (8)$$

$$\hat{c}_p = (\log_2 e) \cdot \sum_{j=j_1}^{j_2} w_j \hat{C}(2^j, p), \quad (9)$$

where $S(2^j, q)$ can be read as the standard sample moment estimator and $\hat{C}(2^j, p)$ refers to the standard cumulant estimator of order p for $\ln L_X(j, k)$ (or $\ln |d_X(j, k)|$). The weights w_j have to satisfy the usual constraints $\sum_{j=j_1}^{j_2} w_j \equiv 1$ and $\sum_{j=j_1}^{j_2} j w_j \equiv 0$ and reflect the confidence granted to each $\log_2 S(j, q)$ or $\hat{C}(j, p)$.

Estimates based on Leaders or wavelet coefficients are denoted as $\hat{\zeta}^L(q)$, \hat{c}_p^L and $\hat{\zeta}^d(q)$, \hat{c}_p^d , respectively.

4. TIME-SCALE BLOCK BOOTSTRAP

In order to design confidence limits and tests for $\hat{\zeta}(q)$ and \hat{c}_p , we extend here the line of work in [3], proposing a time block bootstrap procedure in the wavelet domain, by applying a time-scale block bootstrap procedure to either the d_X or to the L_X . This specific choice of block construction is expected to ensure that both the time

and the scale dependencies of the coefficients are reproduced by the bootstrap resamples (cf. [9, 10] for details on resampling methods).

Time-scale block resampling. Let k denote the time index of the sampled signal $X[k] = X(kT_s)$ with sampling period T_s and j', k' the indices of the d_X and L_X of X . Then, the time-scale blocks are constructed as stripes of time length $2l$ in the time-scale plane, overlapping in time :

$$B_k = \{(j', k') | k \leq k' 2^{j'} \leq k + 2l, 1 \leq j' \leq j_{max}\}, 1 \leq k \leq n.$$

Each of the bootstrap resamples $L_X^{*(r)}$, $r = 1, \dots, R$ is obtained by first drawing, randomly, and with replacement $B = \lceil \frac{n}{2l} \rceil$ such time-scale blocks $B_{k(b)}$, $b = 1, \dots, B$, from the available blocks and then concatenating in time the time-circularized coefficients, $k' 2^{j'} > n : L_X(j', k') = L_X(j', k' - n/2^{j'})$, which belong to each of the drawn blocks:

$$L_X^{*(r)} = \{(L_X(j', k') | (j', k') \in B_{k(b)}), b = 1, \dots, B\}.$$

The procedure is illustrated in Fig. 2. The bootstrap resamples $\hat{\zeta}(q)^{*(r)}$ and $\hat{c}_p^{*(r)}$ are finally obtained by plugging the $L_X^{*(r)}$ into Eqs. (6) and (7) and hence into Eqs. (8) and (9).

Bootstrap variance estimation and confidence limits. The bootstrap estimation of the variance of $\hat{\Theta} \in \{\hat{\zeta}(q), \hat{c}_p\}$ is obtained as the sample variance over the R bootstrap resamples $\hat{\Theta}^*$, $\hat{\sigma}_{\hat{\Theta}}^{*2} = \widehat{\text{Var}}^* \hat{\Theta}^*$. The classical equi-tailed $(1-\alpha)$ percentile confidence limit for the parameter $\Theta \in \{\zeta(q), c_p\}$ is given by $\widehat{\text{CI}}_{\Theta} = [\hat{\Theta}_{\frac{\alpha}{2}}^*, \hat{\Theta}_{(1-\frac{\alpha}{2})}^*]$, where $\hat{\Theta}_{\alpha}^*$ is the α quantile of the empirical distribution of $\hat{\Theta}^*$.

Bootstrap hypothesis tests. We want to test $H_0 : c_p = c_{p,0}$ against the double sided alternative $c_p \neq c_{p,0}$, using the test statistic $\hat{t} = \hat{c}_p - c_{p,0}$. The bootstrap estimation of the null distribution of \hat{t} is obtained as the empirical distribution of $\{\hat{t}^{*(r)} = \hat{c}_p^{*(r)} - \hat{c}_p\}_{r=1}^R$ with α quantiles \hat{t}_{α}^* . The significance $(1-\alpha)$ percentile bootstrap test d_{α} rejects H_0 if \hat{t} is outside of the acceptance region $\widehat{\mathcal{T}}_{(1-\alpha)} = [-\hat{t}_{(1-\frac{\alpha}{2})}^*, \hat{t}_{\frac{\alpha}{2}}^*]$, i.e., $d_{\alpha} = 1$ if $\hat{t} \notin \widehat{\mathcal{T}}_{(1-\alpha)}$, and $d_{\alpha} = 0$ otherwise.

5. RESULTS

Experimental set-up. In order to evaluate the performance of the proposed estimation and test procedures, we apply them to a large number N_{MC} of realizations of length n of FBM and ROS. The simulation parameters are set to $N_{MC} = 500$, $n = 2^{15}$ and $H = 0.7$ for the processes, $N_{\psi} = 3$ and $[j_1, j_2] = [3, 8]$ for estimation, and $R = 199$ and $l = 2^{10}$ for the bootstrap.

Performance evaluation. The estimation performance are quantified through the bias $\gamma_{\hat{\Theta}} = \widehat{\mathbb{E}}_{N_{MC}} \hat{\Theta} - \Theta$, standard deviation $\sigma_{\hat{\Theta}} = \sqrt{\widehat{\text{Var}}_{N_{MC}} \hat{\Theta}}$ and (root) mean squared error $\text{MSE}_{\hat{\Theta}} = \sqrt{\sigma_{\hat{\Theta}}^2 + \gamma_{\hat{\Theta}}^2}$, $\widehat{\mathbb{E}}_{N_{MC}}$ and $\widehat{\text{Var}}_{N_{MC}}$ stand for sample mean and sample variance estimates from N_{MC} independent realizations.

The standard deviation $\sigma_{\hat{\Theta}}$ is compared to its bootstrap counterpart $\sigma_{\hat{\Theta}}^* = \sqrt{\widehat{\mathbb{E}}_{N_{MC}} \hat{\sigma}_{\hat{\Theta}}^{*2}}$. The reliability of the bootstrap based confidence intervals is evaluated through their bias-corrected empirical coverage: $\mathcal{C}_{\Theta, bc} = \widehat{\mathbb{E}}_{N_{MC}} \mathbb{I}\{\Theta + \gamma_{\hat{\Theta}} \in \widehat{\text{CI}}_{\Theta}\}$, where $\mathbb{I}\{\cdot\}$ is the indicator function of the event $\{\cdot\}$. The rejection rates of the tests H_0 :

FBM		d_X			L_X		
Θ		bias	std	mse	bias	std	mse
$\zeta(-2)$	-1.400	-0.995	1.171	1.537	0.033	0.018	0.038
$\zeta(-1)$	-0.700	-0.107	0.323	0.340	0.014	0.009	0.017
$\zeta(-0.5)$	-0.350	-0.001	0.029	0.029	0.006	0.004	0.008
$\zeta(0.5)$	0.350	-0.001	0.007	0.007	-0.005	0.005	0.007
$\zeta(1)$	0.700	-0.002	0.012	0.012	-0.009	0.009	0.013
$\zeta(2)$	1.400	-0.004	0.023	0.023	-0.014	0.020	0.024
c_1	0.700	-0.002	0.017	0.017	-0.011	0.009	0.014
c_2	0.000	-0.001	0.042	0.042	0.005	0.003	0.006
c_3	0.000	0.008	0.217	0.217	-0.001	0.001	0.002

ROS		d_X			L_X		
Θ		bias	std	mse	bias	std	mse
$\zeta(-2)$	-1.400	-1.002	1.101	1.489	0.026	0.035	0.043
$\zeta(-1)$	-0.700	-0.114	0.266	0.289	0.013	0.018	0.022
$\zeta(-0.5)$	-0.350	-0.004	0.023	0.023	0.006	0.009	0.011
$\zeta(0.5)$	0.350	0.001	0.011	0.011	-0.006	0.010	0.012
$\zeta(1)$	0.700	0.001	0.023	0.023	-0.011	0.022	0.024
$\zeta(2)$	1.400	-0.004	0.052	0.052	-0.021	0.050	0.054
c_1	0.700	0.004	0.024	0.024	-0.012	0.019	0.023
c_2	0.000	-0.006	0.043	0.044	0.002	0.009	0.009
c_3	0.000	0.006	0.211	0.211	0.001	0.006	0.006

Table 1: Statistical performance for wavelet coefficient (left) and Leader (right) based estimation procedures, for FBM (top) and ROS (bottom). Multifractal parameters are correctly estimated for both processes. The Leader based estimates for c_2 and c_3 have significantly lower MSEs than the coefficient based ones.

$$c_p = c_{p,0} \text{ are estimated as } \hat{\beta}(\alpha, c_{p,0}, c_p) = \hat{\mathbb{E}}_{NMC} \left\{ \hat{d}_\alpha | c_{p,0}, c_p \right\}.$$

Performance of estimation. Table 1 shows that for $q < 0$, wavelet coefficient based estimations of $\zeta(q)$ exhibit large biases and MSEs and hence are not meaningful, whereas their Leader based counterparts perform satisfactorily well for both FBM and ROS. For $q > 0$, both the coefficient and the Leader based estimations are satisfactory for both FBM and ROS. Note that in this situation, the Leader based estimations are more biased but have smaller standard deviations, resulting in MSEs similar to those of coefficients based ones.

Moreover, we note that for the estimation of the log-cumulants c_p , Leader based estimations have smaller standard deviations and MSEs - and significantly so for $c_p, p = 2, 3$, the key attributes in discriminating mono- vs. multi-fractal models. The gains in MSEs are of approximately 1 and 2 orders of magnitude for c_2 and c_3 , respectively. Together with the estimation performance for $\zeta(q)$, when $q < 0$, these gains in MSE indicate clearly that Leaders are to be preferred over wavelet coefficients for investigating the precise multifractal properties of the data under analysis.

We further observe that the non Gaussian nature of ROS results in larger standard deviations and MSEs - up to a factor 2 as compared to Gaussian FBM - for both coefficient and Leader based estimations. Nevertheless, the estimation performance remains satisfactory for the non Gaussian process. In particular, we note that the Leader based estimations of c_p remain excellent. The non Gaussianity of the wavelet coefficients of ROS is illustrated in Fig. 1.

Performance of bootstrap confidence intervals. Fig. 3 illustrates the relevance of the bootstrap distribution estimations for \hat{c}_2 by comparing the Monte Carlo empirical probability density functions of $(\hat{c}_2 - \hat{\mathbb{E}}_{NMC} \hat{c}_2) / \sigma_{\hat{c}_2}$ to the empirical distributions $(\hat{c}_2^* - \hat{\mathbb{E}}^* \hat{c}_2^*) / \sigma_{\hat{c}_2^*}$

FBM		time-scale block					time block	
Θ		d_X			L_X			L_X
		$\sigma_{\hat{\Theta}}$	$\sigma_{\hat{\Theta}^*}$	$\mathcal{C}_{\Theta,bc}$	$\sigma_{\hat{\Theta}}$	$\sigma_{\hat{\Theta}^*}$	$\mathcal{C}_{\Theta,bc}$	$\sigma_{\hat{\Theta}^*}$
$\zeta(-2)$		1.171	0.718	82.0	0.018	0.017	91.6	0.014
$\zeta(-1)$		0.323	0.200	88.6	0.009	0.008	91.4	0.007
$\zeta(-0.5)$		0.029	0.023	91.0	0.004	0.004	91.6	0.003
$\zeta(0.5)$		0.007	0.006	94.2	0.005	0.004	93.2	0.004
$\zeta(1)$		0.012	0.012	93.4	0.009	0.009	94.0	0.007
$\zeta(2)$		0.023	0.023	94.4	0.020	0.019	94.6	0.016
c_1		0.017	0.016	94.2	0.009	0.009	92.2	0.007
c_2		0.042	0.040	91.8	0.003	0.003	93.6	0.003
c_3		0.217	0.200	90.8	0.001	0.001	94.0	0.001

ROS		time-scale block					time block	
Θ		d_X			L_X			L_X
		$\sigma_{\hat{\Theta}}$	$\sigma_{\hat{\Theta}^*}$	$\mathcal{C}_{\Theta,bc}$	$\sigma_{\hat{\Theta}}$	$\sigma_{\hat{\Theta}^*}$	$\mathcal{C}_{\Theta,bc}$	$\sigma_{\hat{\Theta}^*}$
$\zeta(-2)$		1.101	0.703	81.8	0.035	0.026	88.0	0.020
$\zeta(-1)$		0.266	0.181	89.4	0.018	0.013	86.8	0.010
$\zeta(-0.5)$		0.023	0.020	90.8	0.009	0.007	87.0	0.005
$\zeta(0.5)$		0.011	0.008	84.0	0.010	0.007	85.4	0.006
$\zeta(1)$		0.023	0.017	84.4	0.022	0.015	85.2	0.012
$\zeta(2)$		0.052	0.037	84.0	0.050	0.036	83.8	0.030
c_1		0.024	0.019	88.8	0.019	0.014	85.8	0.011
c_2		0.043	0.041	93.2	0.009	0.008	88.8	0.006
c_3		0.211	0.196	91.0	0.006	0.006	87.6	0.005

Table 2: Performance of time-scale block bootstrap for wavelet coefficient (left) and Leader (middle) based estimation, for FBM (top) and ROS (bottom). Time-scale block bootstrap estimations are excellent for FBM, and remain satisfactory for ROS. The ordinary time block bootstrap estimations for L_X (right) are worse than time-scale block bootstrap estimations. The targeted coverage is $\mathcal{C}^{\Theta,bc} = 95\%$.

estimated by bootstrap (shown are the Monte Carlo median and upper and lower 5% quantiles of these bootstrap distributions). We observe that for both coefficients and Leaders and for both FBM and ROS, the bootstrap distributions closely reproduce the Monte Carlo distributions of the estimates.

Table 2 (columns 2 to 7) shows that for FBM, bootstrap standard deviation estimations for all \hat{c}_p and $\zeta(q)$, $q > 0$ and $q < 0$ are excellent (under-estimation by less than 5%). However, it also indicates that bootstrap standard deviation estimations are more difficult for the non Gaussian ROS process, with under-estimations of up to 25% for \hat{c}_1 and $\zeta(q)$. Nevertheless, the bootstrap standard deviation estimations for \hat{c}_2 and \hat{c}_3 , the key attributes in discriminating mono- vs. multi-fractal models, are still excellent for ROS. Similar observations are obtained for (bias corrected) confidence interval coverages: They are excellent for all estimates for FBM and for \hat{c}_2, \hat{c}_3 for ROS, and remain satisfactory for $\hat{c}_1, \zeta(q)$ for ROS.

Time block vs. Time-scale block bootstrap. For wavelet coefficients, the time block and the time-scale block procedures perform equivalently well for both processes (results not reported here for space reasons). For wavelet Leaders, Table 2 (columns 8 and 9) show the standard deviation and confidence interval estimation performance of an ordinary time block bootstrap (cf. [3, 4]). It is observed that the performance of the time block bootstrap are inferior to those of the time-scale block bootstrap proposed in the present work. Whereas the difference in performance remains small for FBM, it becomes more significant for ROS. Such observations lead us to the conclusion that the non-linear construction of wavelet Lead-

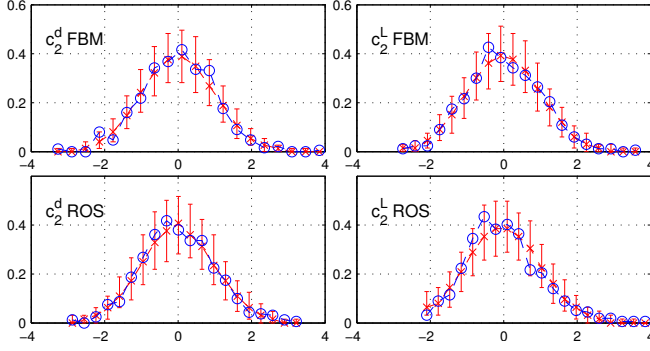


Fig. 3: Centered and normalized empirical probability density functions of \hat{c}_2 , obtained from Monte Carlo simulation (blue dashed line and circles) and bootstrap (red solid line with crosses and error bars for median and 5% quantiles, respectively, obtained from Monte Carlo simulation) for FBM (top) and Rosenblatt (bottom) with wavelet coefficients (left) and Leaders (right).

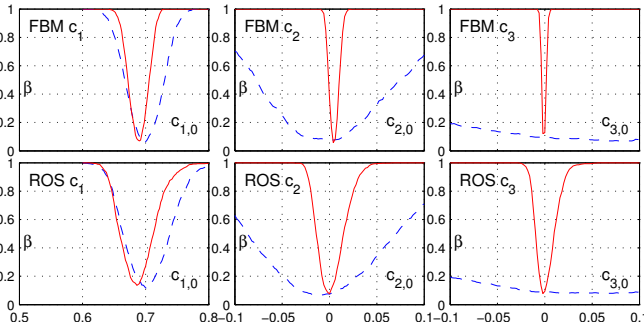


Fig. 4: Rejection rate $\hat{\beta}$ of bootstrap tests for FBM (top) and Rosenblatt (bottom) for coefficients (blue dashed) and Leaders (red solid). The true values c_p are $(c_1, c_2, c_3) = (0.7, 0, 0)$, the null values $c_{p,0}$ are given by the abscissa. It is significantly more difficult to reject a MF hypothesis when data actually are non Gaussian ROS than when they are Gaussian FBM. The target significance of the tests is set to $\alpha = 5\%$.

ers causes an increase of inter-scale dependencies as compared to wavelet coefficients. This justifies the recourse to time-scale block bootstrap procedures when basing scaling (or multifractal) analysis on wavelet Leaders.

Performance of bootstrap tests. Because of the duality of confidence interval and test acceptance region constructions, the significance of the proposed bootstrap tests (in %) for a targeted significance of 5% can be read from Table 2 as $100 - C_{\Theta, (bc)}$. Thus, the conclusion - performance are excellent for FBM and remain satisfactory for ROS - applies as well to test significance.

Rejection rates (powers) $\hat{\beta}$ of tests $H_0 : c_p = c_{p,0}$ are shown in Fig. 4 for FBM (top row) and ROS (bottom row) for both Leaders (solid red lines) and coefficients (dashed blue lines). The powers $\hat{\beta}$ of the tests equal the probabilities of rejecting $c_{p,0}$ (given by the abscissa) when the alternatives c_p (i.e. $\{c_1, c_2, c_3\} = \{0.7, 0, 0\}$) are true. Thus, for $p = 2, 3$, $\hat{\beta}$ quantifies the capability of the test to reject a multifractal hypothesis when the data are self-similar FBM or ROS. We observe that Leader based tests display larger power than coef-

ficient based ones, and significantly so for the parameters c_2 and c_3 which discriminate mono- from multi-fractality. However, switching from Gaussian FBM to non Gaussian ROS results in a non negligible decrease in test powers. Therefore, it is significantly more difficult to reject a multifractal hypothesis when the data are non Gaussian H -sssi processes than when they are Gaussian H -sssi.

6. CONCLUSIONS

In the present contribution, the following results were obtained. First, multifractal parameter estimations work satisfactorily both for Gaussian and non Gaussian H -sssi processes. Wavelet Leader based estimation generally show better performance than those based on coefficients, the difference being significant for the estimation of the parameters enabling discrimination between mono- and multi-fractal models, namely the c_p s and the $\zeta(q)$, when $q < 0$. Along the same line, the Leader based tests are consistently and significantly more powerful. Second, the proposed time-scale block bootstrap procedure provides reliable confidence intervals and test decisions for both Gaussian and non Gaussian H -sssi processes. It is demonstrated to be preferable to time block bootstrap procedures when the multi-resolution quantities display non negligible inter-scale dependencies, as is the case for Leaders. Third, it is observed that, irrespectively of the efficiency of bootstrap, the non Gaussian nature of ROS causes a significant increase in confidence interval sizes and decrease in test powers, and hence increased difficulties in discriminating between multifractality and non Gaussian self-similarity.

7. REFERENCES

- [1] S. Jaffard, "Wavelet techniques in multifractal analysis," in *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot*, M. Lapidus et M. van Frankenhuysen Eds., *Proceedings of Symposia in Pure Mathematics*. 2004, vol. 72(2), pp. 91–152, AMS.
- [2] R. H. Riedi, "Multifractal processes," in: *Theory and applications of long range dependence*, eds. Doukhan, Oppenheim and Taqqu, Birkhäuser, 2003.
- [3] H. Wendt and P. Abry, "Multifractality tests using bootstrapped wavelet leaders," *IEEE Trans. on Sig. Proc.*, to appear, 2007.
- [4] H. Wendt, P. Abry, and S. Jaffard, "Bootstrap for empirical multifractal analysis with application to hydrodynamic turbulence," *IEEE Sig. Proc. Mag.*, to appear, 2007.
- [5] M. Rosenblatt, "Independence and dependence," in *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability*, University of Carolina Press, Berkeley, CA, 1961, vol. 2, pp. 431–443.
- [6] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. on Info. Theory*, vol. IT-38, no. 2, pp. 910–917, 1992.
- [7] G. Samorodnitsky and M. Taqqu, *Stable non-Gaussian random processes*, Chapman and Hall, New York, 1994.
- [8] P. Abry and V. Pipiras, "Wavelet-based synthesis of the rosenblatt process," *Signal Processing*, vol. 86, no. 9, pp. 2326–2339, 2006.
- [9] B. Efron, *The Jackknife, the Bootstrap, and Other Resampling Plans*, SIAM, 1982.
- [10] S.N. Lahiri, *Resampling Methods for Dependent Data*, Springer, New York, 2003.